

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 08-340348

(43)Date of publication of application : 24.12.1996

(51)Int.Cl.

H04L 12/56

(21)Application number : 07-147550

(71)Applicant : KOFU NIPPON DENKI KK

(22)Date of filing : 14.06.1995

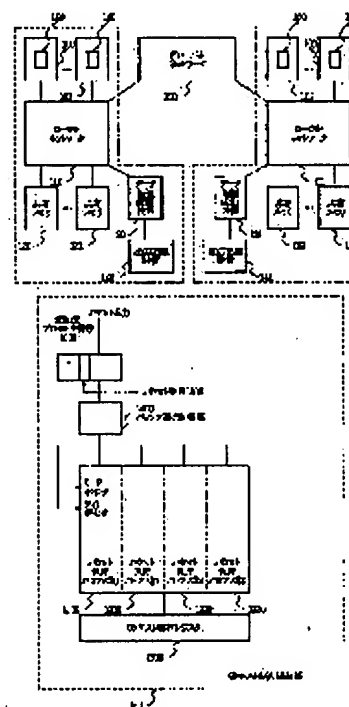
(72)Inventor : ODA MINORU

(54) INFORMATION PROCESSING SYSTEM

(57)Abstract:

PURPOSE: To eliminate the need for exclusive control between processors by arranging packets in mixture in plural processors in a buffer of a data transfer processing unit while using a processor number in each packet as a read pointer and a write pointer.

CONSTITUTION: A request publication control section 160 in arithmetic processors 100 to 103 adds a start source processor number 1000 and a packet number 1100 to a series of instruction packages sent from the processors 100 to 103. Then a buffer to be written based on the number 1100 is selected in data transfer processing units 130, 131, a buffer selection control section 1400 reads data from the buffer by using the start source processor number as a read pointer by the arrival of a final packet. The control section 1400 is controlled by using the start source processor as the read pointer and the write pointer and the data read from the buffer are stored in a request arrangement register 1300 when the final packet reaches. Then a request is started and packets are arranged in the buffer.



LEGAL STATUS

[Date of request for examination] 14.06.1995

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 2710587

[Date of registration] 24.10.1997

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

* NOTICES *

JPO and NCIPI are not responsible for any damages caused by the use of this translation.

1. This document has been translated by computer. So the translation may not reflect the original precisely.
2. **** shows the word which can not be translated.
3. In the drawings, any words are not translated.

CLAIMS

[Claim(s)]

[Claim 1] Two or more processors, the network to which said interprocessor is connected, and the shared memory linked to said network, It consists of data transfer processors which connect with said network and connect extended storage. A means to divide into two or more packets ***** which directs the data transfer between the shared memory to a data transfer processor, and extended storage to a processor, In the information processing system equipped with a means to send out said packet from a processor as a certain requests of a series of to a data transfer processor via the network which has a sequence guarantee function A means to add a processor number to each packet sent out from a processor, Information processing system characterized by having a means to align on a buffer the packet intermingled from two or more processors considering the processor number in a packet as a lead pointer and a light pointer in a data transfer processor.

[Claim 2] Two or more processors, the network to which interprocessor is connected, and the shared memory linked to a network, It consists of data transfer processors which connect with a network and connect extended storage. A means to have the instruction which directs the data transfer between the shared memory to a data transfer processor, and extended storage to a processor, and to divide said instruction into two or more packets, In the information processing system equipped with a means to send out said packet from a processor as a certain requests of a series of to a data transfer processor via the network which has a sequence guarantee function A means to add a packet number to each packet of the request of a certain single string sent out from a processor, and to detect the last packet in a data transfer processor, Information processing system characterized by having a means to perform control which reads data from said buffer by the last packet, and the disappearance error detection means of the packet which checks that all packets have arrived at the time of the last packet detection.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DETAILED DESCRIPTION

[Detailed Description of the Invention]

[0001]

[Industrial Application] About the information processing system which performs data transfer instruction between the shared memories and extended storage in a multiprocessor system, especially, this invention decomposes a series of requests into two or more packets from a processor, and relates to the information processing system sent out to a data transfer processor.

[0002]

[Description of the Prior Art] As the conventional information processing system is shown in drawing 4, drawing 5, drawing 6, and drawing 7. It has the lock byte 310 for exclusive control of request sending out in a data transfer processor on a shared memory 120. Each processors 100-103 A request issue means 200 to decode the instruction (henceforth, data transfer instruction) which directs the data transfer between shared memories 120-123 and extended storage 140 and 141, and to divide this into two or more packets as a series of requests. A packet transmitting means 210 to transmit a packet to a data transfer processor via a network. A lock means 220 to lock or unlock the lock byte on a shared memory. It consists of lock success decision means 300 to judge a lock success or un-succeeding. Each of the data transfer processor 130,131 The packet alignment means 400 which the packet from a processor is stored, and two or more packets are aligned, and is considered as one request. A request decode means 410 to decode the request from a processor. A share memory access means 430 to perform read-out from a shared memory, and the writing to a shared memory. An extended storage access means 440 to perform read-out and the writing from extended storage. It consists of a transfer control means 420 which controls the share memory access means 430 and the extended storage access means 440 by directions of the decode means 410, and a notice means 450 of termination to judge data transfer termination and to report termination to the processor of a requiring agency.

[0003] Next, actuation of the information processing system shown in drawing 4 thru/or drawing 7 is explained. The case where a processor 100 publishes data transfer instruction between a shared memory 120 and extended storage 140 is considered. By the request issue means 200, the instruction published by the processor 100 is decoded with data transfer instruction. Then, the operation processor 100 locks the lock byte on a shared memory with the lock means 220, in order [to the data transfer processor 130] to acquire the right of issue of a request. When it is judged by the lock success decision means 300 here that the lock was unsuccessful, it reruns until it succeeds. When it is judged that the lock was successful, the operation processor 100 divides data transfer instruction into two or more packets with the request issue means 200, and transmits this packet to the data transfer processor 130 via a local network 110 with the packet sending-out means 210. The flag of a processor number and the last packet is added to the last packet of a series of requests by the request issue means 200 here. The data transfer processor 130 receives the packet from a processor 100, and stores a packet in the received sequence at the request alignment means 400. The request alignment means 400 notifies the completion of request reception for the packet to which the last packet flag was added to the request decode means 410 at the time of reception. The request decode means 410 reads and decodes the request from the processor 100 stored in the request alignment means 400. Since a request is the transfer to extended storage 140 from a shared memory 120, the transfer control means 420 takes out shared memory read-out directions from the request decode means 410 for the read-out starting address of a shared memory 120, and transfer length to the share memory access means 430 for every reception and data transfer batch. The share memory access means 430 notifies read-out completion to the transfer control means 420 while passing the read data to the extended storage access means 440. Subsequently, the transfer control means 420 takes out the write-in directions to extended storage from the request decode means 410 for the write-in starting address of extended storage 130 to the extended storage access means 440 for every reception and data transfer batch. The extended storage access means 440 writes the shared memory read-out data from the share memory access means 430 in extended storage 140, and notifies write-in completion to the transfer control means 420. The transfer control means 420 judges data transfer termination by this notice, and reports data transfer termination to the notice means 450 of termination. The notice means 450 of termination notifies termination of data transfer instruction for the number of a processor 100 to reception and the requiring agency processor 100 by local network 110 course from the request decode means 420. A processor 100 receives this notice and unlocks the lock byte 310 on a shared memory with the lock means 220.

[0004]

[Problem(s) to be Solved by the Invention] With this conventional information processing system, since exclusive control (a lock byte's control) was needed by interprocessor at the time of two or more packet sending out of a series of requests from each processor to a data transfer processor in order to align two or more packets, the processor had the fault from which it becomes complicated processing an overhead greatly by the data transfer between a shared memory and extended storage. Moreover, since management of two or more packets sent out as a series of requests was not able to be performed, when a processor broke down during a series of request sending out, there was also a fault which serves as a system down, without the ability performing degeneration operation by separation of only this processor.

[0005]

[Means for Solving the Problem] The network where this invention connects said interprocessor to two or more processors. It consists of data transfer processors which connect with the shared memory linked to said network in said network, and connect extended storage. A means to divide into two or more packets ***** which directs the data transfer between the shared memory to a data transfer processor, and extended storage to a processor. In the information processing system equipped with a means to send out said packet from a processor as a certain requests of a series of to a data transfer processor via the network which has a sequence guarantee function A means to add a processor number to each packet sent out from a processor. It is characterized by having a means to align on a buffer the packet

intermingled from two or more processors considering the processor number in a packet as a lead pointer and a light pointer in a data transfer processor.

[0006] The network where this invention connects interprocessor to two or more processors, It consists of data transfer processors which connect with the shared memory linked to a network in a network, and connect extended storage. A means to have the instruction which directs the data transfer between the shared memory to a data transfer processor, and extended storage to a processor, and to divide said instruction into two or more packets, In the information processing system equipped with a means to send out said packet from a processor as a certain requests of a series of to a data transfer processor via the network which has a sequence guarantee function A means to add a packet number to each packet of the request of a certain single string sent out from a processor, and to detect the last packet in a data transfer processor, It is characterized by having a means to perform control which reads data from said buffer by the last packet, and the disappearance error detection means of the packet which checks that all packets have arrived at the time of the last packet detection.

[0007]

[Example] Next, this invention is explained with reference to a drawing.

[0008] Drawing 1 is the block diagram showing the example of the information processing system by this invention. The information processing system shown in drawing 1 The 1st cluster which has two or more processing units 100,101, two or more shared memories 120,121, the data transfer processor 130, the local network 110 that interconnects them, and the extended storage 140 connected to the data transfer processor 130, The 2nd cluster which has two or more processing units 102,104, two or more shared memories 122,123, the data transfer processor 131, the local network 111 that interconnects them, and the extended storage 141 connected to the data transfer processor 131, The global network 150 which connects between the said 1st and 2nd cluster mutually is provided.

[0009] drawing 1 -- being, processing units 100-103 have the function in which the directions from an operator perform instruction processing. The local network 110 (111) has the function to perform access to the shared memory (120-123) made into the purpose with the instruction from an operation processor (100-103), the function which sends out an instruction to perform data transfer between a shared memory (120-123) and extended storage (140-141) to a data transfer processor (130,131), and the function which sends out the instruction to other clusters to global network 150. A shared memory (120-123) is written through a local network (110,111). The data transfer processor (130,131) has the function which aligns the instruction which consists of two or more packets, the function to decode an instruction, and the function in which the decoded instruction performs data transfer between a shared memory (120-123) and extended storage (140,141). Extended storage (140,141) is written by the data transfer processor (130,131). Global network 150 is sent from a local network (101-103), it targets the instruction to a cluster, and also it has the function sent out to the local network (101-103) of a cluster.

[0010] As shown in drawing 2, in the 1st example of this invention the request issue control section 160 in an operation processor (100-103) The starting agency processor number 1000 and the packet number 1100 are added to the packet of a series of instructions sent out from an operation processor (100-103). The buffer which should be written in by the packet number in a data transfer processor (130,131) is chosen, and it has the buffer selection-control section 1400 which leads the data from a buffer by using a starting agency processor number as a lead pointer by arrival of the last packet. This buffer selection-control section 1400 is equipped with the packet maintenance buffers 1200, 1210, 1220, and 1230 controlled considering said starting agency processor number as a lead pointer and a light pointer the number of packets (this example four pieces). The data led by arrival of the last packet from the buffer are stored in the request alignment register 1300, and a request is actually started after that. Moreover, the sequence between packets shall be guaranteed in the network (a local network, global network).

[0011] In the 2nd example of this invention, as shown in drawing 3, it has the request arrival flags 2000-2110 in the buffers of each, and has the error detection section 2200 which detects the disappearance error of the packet which checks that all packets have arrived in a data transfer processor (130,131) at the time of the last packet detection of a request with a request arrival flag.

[0012] Next, actuation of the 1st example of this invention is explained with reference to drawing 1 and drawing 2.

[0013] Suppose that data transfer instruction which performs data transfer to extended storage 140 was published from a shared memory 120 to a processing unit 100 to the data transfer processor 130. Suppose that data transfer instruction which performs data transfer to it and coincidence from a processing unit 103 to extended storage 140 was published. As for data transfer instruction, the packet number to which one instruction expresses packet sequence as the starting agency processor number which is divided into four packets and expresses a starting agency processor to each packet shall be added. First, the packet number "1000" of the instruction published from the processing unit 100 and the packet number "1030" of the instruction published from the processing unit 103 are inputted into a local network 110, and are arbitrated by the local network 110, a packet number "1000" is sent out first to the data transfer processor 130, and then a packet number "1030" is inputted into the data transfer processor 130. In the data transfer processor 130, the starting agency processor number "100" of the packet number "1000" of the sent request determines the light pointer number "000" of a buffer, and the packet number "0" of a request determines the writing to a buffer number "0." Moreover, the starting agency processor number of the packet number "1030" of the request sent to the degree determines the light pointer number "002" of a buffer, and the writing of buffer #0 is ****(ed) by packet number #0 of a request. Next, the packet number "1001" of the instruction published from the processing unit 100 is arbitrated by the local network 110 like a packet number "1000", and is sent out to the data transfer processor 130. Moreover, the packet number "1031" of the instruction published from operation Prime Minister equipment 103 is similarly arbitrated by the local network 110, and it is sent out to the degree of the packet number "1001" of an instruction published from the processing unit 100 to the data transfer processor 130. In the data transfer processor 130, the starting agency processor number "100" of the packet number "1001" of the sent request determines the light pointer number "000" of a buffer, and the packet number "1" of a request determines the writing to a buffer number "1." Moreover, the starting agency processor number "103" of the packet number "1031" of the request sent to the degree determines the light pointer number "002" of a buffer, and the packet number "1" of a request determines the writing to a buffer number "1." Similarly, a packet number "1002" is sent out from a processing unit 100, and is written in the light pointer number "000" of the buffer number in a data transfer processor "2." Moreover, a packet number "1032" is sent out from a processing unit 103, and is written in the light pointer number "103" of the buffer number in a data transfer processor "2." The packet number "1003" of the instruction finally published from the processing unit 100 is arbitrated by the local network 110, and is sent out to the data transfer processor 130. In the data transfer processor 130, the starting agency processor number "100" of the sent request determines the light pointer number "000" of a buffer, and the packet number "3" of a request performs the writing to a buffer number "3." Moreover, a packet number "3" determines the lead actuation from the buffer in degree timing, and a lead pointer number "000" is determined from a starting agency processor number "100." The lead data from buffer number "0"- "3" are made to input into a flip-flop, and request alignment of two or more packets is made to complete. The packet number "1033" of the instruction similarly sent out from the processing unit 103 is arbitrated by the local network 110, and is sent out to the degree

of the packet number "1003" of the request sent out from the processing unit 100 to the data transfer processor 130. In a data transfer processor, the starting agency processor number "103" of the sent request determines the light pointer number "002" of a buffer, and the packet number "3" of a request performs the writing to a buffer number "3." Moreover, a packet number "3" determines the lead actuation from the buffer in degree timing, determine a lead pointer number "103" from a starting agency processor number "103", the lead data from buffer #0 - a number "3" are made to input into a flip-flop, and alignment of two or more packets is made to complete. In a data transfer processor, these requests are decoded and data transfer between a shared memory and extended storage is performed. If a transfer is completed, it will become completion of data transfer instruction.

[0014] Next, actuation of the 2nd example of this invention is explained with reference to drawing 1 and drawing 3.

[0015] Suppose that data transfer instruction which performs data transfer to extended storage 140 was published from a shared memory 120 to a processing unit 100 to the data transfer processor 130. Moreover, as for data transfer instruction, the packet number to which one instruction expresses packet sequence as the processing unit number which is divided into four packets and expresses a processing unit 100 to each packet shall be added. First, the packet number "0" of the instruction published from the processing unit 100 is sent out through a local network 110 to the data transfer processor 130. The data transfer processor 130 determines the light pointer number "000" of a buffer by the starting agency processor number "100" of the sent request, and determines the writing to a buffer number "0" by the packet number "0" of a request. Moreover, the arrival packet management bit 2000 of the light pointer number "000" of a buffer number "0" is set. Next, the packet number "1" of the instruction published from the processing unit 100 is sent out through a local network like a packet number "0" to the data transfer processor 130. The data transfer processor 130 determines the light pointer number "000" of a buffer by the starting agency processor number "100" of the sent request, and determines the writing to a buffer number "1" by the packet number "1" of a request. Moreover, the arrival packet management bit 2100 of the light pointer number "000" of a buffer number "1" is set. Similarly, a packet number "2" is sent out from a processing unit 100, and is written in the buffer number in a data transfer processor "2." Moreover, the arrival packet management bit 2020 of the light pointer number "000" of a buffer number "2" is set. The packet number "3" of the instruction finally published from the processing unit 100 is sent out through a local network 110 to the data transfer processor 130. The data transfer processor 130 determines the lead pointer number "000" of a buffer by the starting agency processor number number "100" of the sent request. Coincidence is made to input the lead data from buffer number "0" - "2", and the data of a packet number "3" into a flip-flop, and the request of two or more packets is made to complete. Moreover, the management bits 2000-2020 corresponding to the starting agency processor number "000" of an arrival packet management bit are sent out to coincidence at the error detection section, and the error detection section checks that all packets have arrived, when the packet has disappeared, it reports a packet disappearance error, and it performs **** operation by separation of only this processor by the starting agency processor number.

[0016]

[Effect of the Invention] As explained above, according to this invention, at the time of two or more packet sending out in the data transfer processor from each processor, it becomes exclusive control needlessness by interprocessor, and an overhead is lost to a processor by the data transfer between a shared memory and extended storage. Moreover, when management of two or more packets sent out as a series of requests is completed and a processor breaks down during a series of request sending out, degeneration operation by separation of only this processor is attained. Moreover, the buffer of the last packet becomes unnecessary and the amount of hardware can be reduced.

[Translation done.]

* NOTICES *

JPO and NCIPi are not responsible for any damages caused by the use of this translation.

- 1.This document has been translated by computer. So the translation may not reflect the original precisely.
- 2.*** shows the word which can not be translated.
- 3.In the drawings, any words are not translated.

DESCRIPTION OF DRAWINGS

[Brief Description of the Drawings]

[Drawing 1] It is the block diagram of the example of this invention.

[Drawing 2] It is the block diagram of the request alignment control section in the 1st example of this invention.

[Drawing 3] It is the block diagram of the request alignment control section in the 2nd example of this invention.

[Drawing 4] It is the block diagram of the conventional example.

[Drawing 5] It is the block diagram of the shared memory in drawing 5.

[Drawing 6] It is the block diagram of the data transfer processor in drawing 5.

[Drawing 7] It is the block diagram of the operation processor in drawing 5.

[Description of Notations]

100-103 Processing unit

110-111 Local network

120-123 Shared memory

130-131 Data transfer processor

140-141 Extended storage

150 Global Network

160 Request Issue Control Section

200 Request Issue Means

210 Packet Sending-Out Means

220 Lock Means

300 Lock Success Decision Means

310 Lock Byte

400 Request Alignment Means

410 Request Decode Means

420 Transfer Control Means

430 Share Memory Access Means

440 Extended Storage Access Means

450 Notice Means of Termination

1000 Starting Agency Processor Number

1100 Packet Number

1200-1230 Packet maintenance buffer

1300 Request Alignment Register

1400 Packet Maintenance Buffer Selection-Control Section

2000-2080 Arrival packet management bit

2200 Error Detection Section

[Translation done.]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平8-340348

(43) 公開日 平成8年(1996)12月24日

(51) Int.Cl.⁶
H 0 4 L 12/56

識別記号 庁内整理番号
9466-5K

F I
H 0 4 L 11/20

技術表示箇所

1 0 2 Z

審査請求 有 請求項の数 2 O L (全 7 頁)

(21) 出願番号 特願平7-147550

(22) 出願日 平成7年(1995)6月14日

(71) 出願人 000168285

甲府日本電気株式会社
山梨県甲府市大津町1088-3

(72) 発明者 小田 稔

山梨県甲府市大津町1088-3 甲府日本電
気株式会社内

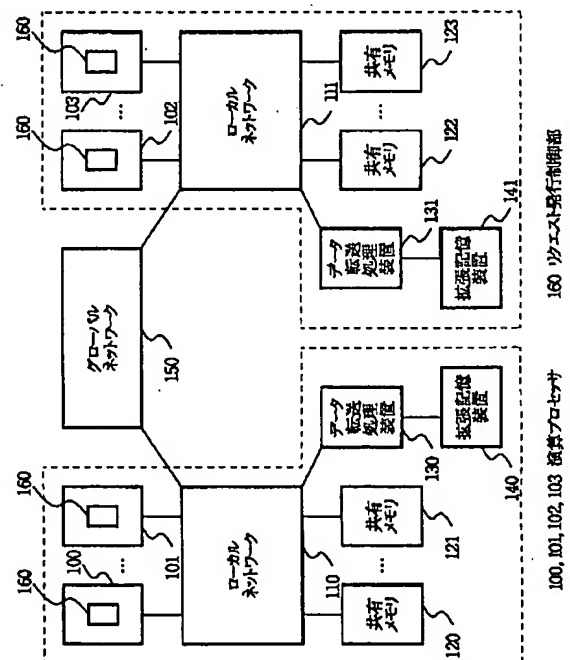
(74) 代理人 弁理士 京本 直樹 (外2名)

(54) 【発明の名称】 情報処理システム

(57) 【要約】

【目的】 高並列スーパーコンピュータにおいて、一連のリクエストとして送出される複数のパケットの管理を簡略化する。

【構成】 高並列スーパーコンピュータの一連のリクエストに起動元プロセッサ番号1000およびパケット番号1100を付加し、前記起動元プロセッサ番号、パケット番号により制御されるバッファ1200~1230 (または1220) を備える。前記バッファにより複数のパケットの管理を簡略化する。



【特許請求の範囲】

【請求項1】 複数のプロセッサと、前記プロセッサ間を結ぶネットワークと、前記ネットワークに接続する共有メモリと、前記ネットワークに接続し拡張記憶装置を接続するデータ転送処理装置から構成され、プロセッサにデータ転送処理装置への共有メモリと拡張記憶装置との間のデータ転送を指示する命令を複数のバケットに分割する手段と、前記バケットをプロセッサから順序保証機能を有するネットワーク経由でデータ転送処理装置に対し、ある一連のリクエストとして送出する手段とを備えた情報処理システムにおいて、

プロセッサから送出される各々のバケットにプロセッサ番号を付加する手段、データ転送処理装置内にバケット内のプロセッサ番号をリードポインタ、ライトポインタとして複数のプロセッサからの混在したバケットをバッファ上で整列させる手段とを備えることを特徴とする情報処理システム。

【請求項2】 複数のプロセッサと、プロセッサ間を結ぶネットワークと、ネットワークに接続する共有メモリと、ネットワークに接続し拡張記憶装置を接続するデータ転送処理装置から構成され、プロセッサにデータ転送処理装置への共有メモリと拡張記憶装置との間のデータ転送を指示する命令を有し、前記命令を複数のバケットに分割する手段と、前記バケットをプロセッサから順序保証機能を有するネットワーク経由でデータ転送処理装置に対し、ある一連のリクエストとして送出する手段を備えた、情報処理システムにおいて、

プロセッサから送出されるある一連のリクエストの各々のバケットにバケット番号を付加し、データ転送処理装置内に最終バケットを検出する手段と、最終バケットにより前記バッファよりデータを読み出す制御を行う手段と、最終バケット検出時、すべてのバケットが到着していることを確認するバケットの消失エラー検出手段を備えることを特徴とする情報処理システム。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、マルチプロセッサシステムにおける共有メモリと拡張記憶装置との間のデータ転送命令を行う情報処理システムに関し、特に、プロセッサから一連のリクエストを複数のバケットに分解し、データ転送処理装置に送出する情報処理システムに関する。

【0002】

【従来の技術】従来の情報処理システムは、

図4、図5、図6、図7に示すように、共有メモリ120上にデータ転送処理装置へのリクエスト送出の排他制御用のロックバイト310を持ち、各プロセッサ100～103は、共有メモリ120～123と拡張記憶装置140、141との間のデータ転送を指示する命令（以後、データ転送命令）を解釈し、これを一連のリクエ

ストとして複数のバケットに分割するリクエスト発行手段200と、ネットワーク経由でデータ転送処理装置にバケットを送信するバケット送信手段210と、共有メモリ上のロックバイトをロックまたはアンロックするロック手段220と、ロック成功あるいは不成功を判断するロック成功判断手段300とから構成され、データ転送処理装置130、131のそれぞれは、プロセッサからのバケットを格納し、複数のバケットを整列させ、ひとつのリクエストとするバケット整列手段400と、プロセッサからのリクエストの解釈を行うリクエスト解釈手段410と、共有メモリからの読み出しおよび共有メモリへの書き込みを行う共有メモリアクセス手段430と、拡張記憶装置からの読み出しおよび書き込みを行う拡張記憶装置アクセス手段440と、解読手段410の指示により共有メモリアクセス手段430と拡張記憶装置アクセス手段440を制御する転送制御手段420と、データ転送終了を判断し、要求元のプロセッサに終了を報告する終了通知手段450とから構成される。

【0003】次に図4乃至図7に示す情報処理システムの動作を説明する。プロセッサ100が共有メモリ120と拡張記憶装置140との間のデータ転送命令を発行した場合を考える。プロセッサ100により発行された命令はリクエスト発行手段200により、データ転送命令と解釈される。その後、演算プロセッサ100は、データ転送処理装置130へのリクエストの発行の権利を得るために、ロック手段220により共有メモリ上のロックバイトのロックを行う。ここでロック成功判断手段300によりロックが不成功であったと判断された場合、成功するまで再実行する。ロックが成功した

と判断された場合、演算プロセッサ100は、データ転送命令をリクエスト発行手段200により複数のバケットに分割し、このバケットをバケット送出手段210によりローカルネットワーク110を経由し、データ転送処理装置130へ送信する。ここで一連のリクエストの最終バケットにはリクエスト発行手段200によりプロセッサ番号と最終バケットのフラグが付加される。データ転送処理装置130は、プロセッサ100からのバケットを受信し、受信した順番にリクエスト整列手段400にバケットを格納する。最終バケットフラグの付加されたバケットを受信時、リクエスト整列手段400はリクエスト受付完了をリクエスト解釈手段410に通知する。リクエスト解釈手段410はリクエスト整列手段400に格納されたプロセッサ100からのリクエストを読み出し解釈する。リクエストが共有メモリ120から拡張記憶装置140への転送であるので、転送制御手段420は、リクエスト解釈手段410から共有メモリ120の読み出し開始アドレス、転送長を受け取り、データ転送処理単位毎に共有メモリアクセス手段430に対して共有メモリ読み出し指示を出す。共有メモリアクセス手段430は、読み出したデータを拡張記憶装置アク

セス手段440に渡すと共に、読み出し完了を転送制御手段420に通知する。次いで、転送制御手段420は、リクエスト解読手段410から拡張記憶装置130の書き込み開始アドレスを受け取り、データ転送処理単位毎に拡張記憶装置アクセス手段440に対して拡張記憶装置への書き込み指示を出す。拡張記憶装置アクセス手段440は、共有メモリアクセス手段430からの共有メモリ読み出しデータを拡張記憶装置140へ書き込み、書き込み完了を転送制御手段420に通知する。転送制御手段420はこの通知によりデータ転送終了を判断し、終了通知手段450にデータ転送終了を報告する。終了通知手段450はリクエスト解読手段420からプロセッサ100の番号を受け取り、要求元プロセッサ100にデータ転送命令の終了をローカルネットワーク110経由で通知する。プロセッサ100は本通知を受け、ロック手段220により共有メモリ上のロックバイト310をアンロックする。

【0004】

【発明が解決しようとする課題】この従来の情報処理システムでは、各プロセッサからデータ転送処理装置への一連のリクエストの複数のパケット送出時、複数のパケットを整列させるためプロセッサ間で排他制御（ロックバイトの制御）が必要となるため、共有メモリと拡張記憶装置との間のデータ転送でプロセッサにオーバーヘッドが大きく、かつ処理が煩雑となる欠点があった。また、一連のリクエストとして送出される複数のパケットの管理ができないため、一連のリクエスト送出中にプロセッサが故障した場合、該プロセッサのみの切り離しによる縮退運転ができずにシステムダウンとなる欠点もあった。

【0005】

【課題を解決するための手段】本発明は、複数のプロセッサと、前記プロセッサ間を結ぶネットワークと、前記ネットワークに接続する共有メモリと、前記ネットワークに接続し拡張記憶装置を接続するデータ転送処理装置から構成され、プロセッサにデータ転送処理装置への共有メモリと拡張記憶装置との間のデータ転送を指示する命令を複数のパケットに分割する手段と、前記パケットをプロセッサから順序保証機能を有するネットワーク経由でデータ転送処理装置に対し、ある一連のリクエストとして送出する手段とを備えた情報処理システムにおいて、プロセッサから送出される各々のパケットにプロセッサ番号を付加する手段、データ転送処理装置内にパケット内のプロセッサ番号をリードポインタ、ライトポインタとして複数のプロセッサからの混在したパケットをバッファ上で整列させる手段とを備えることを特徴とするものである。

【0006】本発明は、複数のプロセッサと、プロセッサ間を結ぶネットワークと、ネットワークに接続する共有メモリと、ネットワークに接続し拡張記憶装置を接続

するデータ転送処理装置から構成され、プロセッサにデータ転送処理装置への共有メモリと拡張記憶装置との間のデータ転送を指示する命令を有し、前記命令を複数のパケットに分割する手段と、前記パケットをプロセッサから順序保証機能を有するネットワーク経由でデータ転送処理装置に対し、ある一連のリクエストとして送出する手段を備えた、情報処理システムにおいて、プロセッサから送出されるある一連のリクエストの各々のパケットにパケット番号を付加し、データ転送処理装置内に最終パケットを検出する手段と、最終パケットにより前記バッファよりデータを読み出す制御を行う手段と、最終パケット検出時、すべてのパケットが到着していることを確認するパケットの消失エラー検出手段を備えることを特徴とするものである。

【0007】

【実施例】次に本発明について、図面を参照して説明する。

【0008】図1は、本発明による情報処理システムの実施例を示す構成図である。図1に示す情報処理システムは、複数の演算処理装置100、101、複数の共有メモリ120、121、データ転送処理装置130、それらを相互接続するローカルネットワーク110、データ転送処理装置130に接続される拡張記憶装置140を有する第1のクラスタと、

複数の演算処理装置102、104、複数の共有メモリ122、123、データ転送処理装置131、それらを相互接続するローカルネットワーク111、データ転送処理装置

131に接続される拡張記憶装置141を有する第2のクラスタと、前記第1、第2のクラスタ間を相互に接続するグローバルネットワーク150とを具備する。

【0009】図1において、演算処理装置100～103は、オペレータからの指示により命令処理を行う機能を有している。ローカルネットワーク110（111）は、演算プロセッサ（100～103）からの命令により目的とする共有メモリ（120～123）へのアクセスを行う機能と、共有メモリ（120～123）と拡張記憶装置（140～141）間のデータ転送を行う目的の命令をデータ転送処理装置（130、131）へ送出する機能と、他クラスタへの命令をグローバルネットワーク150へ送出する機能とを有している。共有メモリ（120～123）はローカルネットワーク（110、111）を介して読み書きされる。データ転送処理装置（130、131）は、複数のパケットからなる命令を整列する機能と、命令を解読する機能と、解読した命令により共有メモリ（120～123）と拡張記憶装置（140、141）間のデータ転送を行う機能とを有している。拡張記憶装置（140、141）はデータ転送処理装置（130、131）により読み書きされる。グローバルネットワーク150は、ローカルネットワーク

(101~103)から送られてくる他クラスタへの命令を、目的とする他クラスタのローカルネットワーク(101~103)へ送出する機能を有している。

【0010】本発明の第1の実施例では、図2に示すように、演算プロセッサ(100~103)内のリクエスト発行制御部160は、演算プロセッサ(100~103)から送出される一連の命令のパケットに起動元プロセッサ番号1000、パケット番号1100を付加し、データ転送処理装置(130、131)内にパケット番号により書き込まれるべきバッファを選択し、最終パケットの到着によりバッファからのデータを、起動元プロセッサ番号をリードポイントとしてリードを行うバッファ選択制御部1400を備える。このバッファ選択制御部1400は、前記起動元プロセッサ番号をリードポイント、ライトポイントとして制御されるパケット保持バッファ1200、1210、1220、1230をパケット数(この実施例では4個)備えている。最終パケットの到着によりバッファからリードされたデータはリクエスト整列レジスタ1300に格納され、その後リクエストが実際に起動される。またパケット間の順序はネットワーク(ローカルネットワーク、グローバルネットワーク)で保証されているものとする。

【0011】本発明の第2の実施例では、図3に示すように、バッファ各々にリクエスト到着フラグ2000~2110を有し、データ転送処理装置

(130、131)内に、リクエストの最終パケット検出時、すべてのパケットが到着していることをリクエスト到着フラグにより確認するパケットの消失エラーを検出するエラー検出部2200を有している。

【0012】次に、

図1、図2を参照して、本発明の第1の実施例の動作を説明する。

【0013】演算処理装置100からデータ転送処理装置130に対して、共有メモリ120から拡張記憶装置140に対してデータ転送を行うデータ転送命令が発行されたとする。それと同時に、演算処理装置103から拡張記憶装置140に対しデータ転送を行うデータ転送命令が発行されたとする。データ転送命令は1つの命令が4つのパケットに分割され、各々のパケットに起動元プロセッサを表す起動元プロセッサ番号と、パケット順序を表すパケット番号とが付加されているものとする。まず、演算処理装置100から発行された命令のパケット番号・1000と、演算処理装置103から発行された命令のパケット番号・1030はローカルネットワーク110に入力され、ローカルネットワーク110で調停され、パケット番号“1000”が最初にデータ転送処理装置130へ送出され、次にパケット番号“1030”がデータ転送処理装置130に入力される。データ転送処理装置130では、送られてきたリクエストのパケット番号“1000”の起動元プロセッサ番号

“100”によりバッファのライトポイント番号“000”を決定し、リクエストのパケット番号“0”によりバッファ番号・0への書き込みを決定する。また次に送られてきたリクエストのパケット番号“1030”の起動元プロセッサ番号によりバッファのライトポイント番号“002”を決定し、リクエストのパケット番号#0によりバッファ#0への書き込みを毛呈する。次に演算処理装置100から発行された命令のパケット番号“1001”はパケット番号“1000”と同様にローカルネットワーク110で調停され、データ転送処理装置130へ送出される。また演算処理装置103から発行された命令のパケット番号“1031”も同様にローカルネットワーク110で調停され、演算処理装置100から発行された命令のパケット番号“1001”の次にデータ転送処理装置130へ送出される。データ転送処理装置130では、送られてきたリクエストのパケット番号“1001”の起動元プロセッサ番号・100によりバッファのライトポイント番号“000”を決定し、リクエストのパケット番号・1によりバッファ番号“1”への書き込みを決定する。また次に送られてきたリクエストのパケット番号“1301”の起動元プロセッサ番号・103によりバッファのライトポイント番号“002”を決定し、リクエストのパケット番号“1”によりバッファ番号“1”への書き込みを決定する。同様にパケット番号“1002”が演算処理装置100から送出されデータ転送処理装置内のバッファ番号“2”のライトポイント番号“000”へ書き込まれる。またパケット番号“1032”が演算処理装置103から送出されデータ転送処理装置内のバッファ番号“2”のライトポイント番号“103”へ書き込まれる。最後に演算処理装置100から発行された命令のパケット番号“1003”はローカルネットワーク110で調停され、データ転送処理装置130へ送出される。データ転送処理装置130では、送られてきたリクエストの起動元プロセッサ番号番号“100”によりバッファのライトポイント番号“000”を決定し、リクエストのパケット番号“3”によりバッファ番号“3”への書き込みを行う。またパケット番号“3”により次タイミングでのバッファからのリード動作を決定し、起動元プロセッサ番号“100”からリードポイント番号“000”を決定する。バッファ番号“0”~“3”からのリードデータをフリップフロップに入力させ複数のパケットのリクエスト整列を完了させる。同様に演算処理装置103から送出された命令のパケット番号“1033”はローカルネットワーク110で調停され、演算処理装置100から送出されたリクエストのパケット番号“1003”の次にデータ転送処理装置130へ送出される。データ転送処理装置では送られてきたリクエストの起動元プロセッサ番号“103”によりバッファのライトポイント番号“002”を決定し、リクエストのパ

ケット番号“3”によりバッファ番号“3”への書き込みを行う。またケット番号“3”により次タイミングでのバッファからのリード動作を決定し、起動元プロセッサ番号“103”からリードポインタ番号“103”を決定し、バッファ#0～番号“3”からのリードデータをフリップフロップに入力させ複数のケットの整列を完了させる。データ転送処理装置では、これらのリクエストを解釈し、共有メモリと拡張記憶装置間のデータ転送を行う。転送が完了すればデータ転送命令の完了となる。

【0014】次に図1、図3を参照して、本発明の第2の実施例の動作を説明する。

【0015】演算処理装置100からデータ転送処理装置130に対して、共有メモリ120から拡張記憶装置140に対してデータ転送を行うデータ転送命令が発行されたとする。またデータ転送命令は1つの命令が4つのケットに分割され、各々のケットに演算処理装置100を表す演算処理装置番号と、ケット順序を表すケット番号とが付加されているものとする。まず、演算処理装置100から発行された命令のケット番号“0”はローカルネットワーク110を介しデータ転送処理装置130へ送出される。データ転送処理装置130は、送られてきたリクエストの起動元プロセッサ番号“100”によりバッファのライトポインタ番号“000”を決定し、リクエストのケット番号“0”によりバッファ番号“0”への書き込みを決定する。また、バッファ番号“0”のライトポインタ番号“000”の到着ケット管理ビット2000をセットする。次に演算処理装置100から発行された命令のケット番号“1”はケット番号“0”と同様に、ローカルネットワークを介しデータ転送処理装置130へ送出される。データ転送処理装置130は、送られてきたリクエストの起動元プロセッサ番号“100”によりバッファのライトポインタ番号“000”を決定し、リクエストのケット番号“1”によりバッファ番号“1”への書き込みを決定する。また、バッファ番号“1”のライトポインタ番号“000”の到着ケット管理ビット2100をセットする。同様にケット番号“2”が演算処理装置100から送出されデータ転送処理装置内のバッファ番号“2”へ書き込まれる。また、バッファ番号“2”のライトポインタ番号“000”の到着ケット管理ビット2020をセットする。最後に演算処理装置100から発行された命令のケット番号“3”はローカルネットワーク110を介しデータ転送処理装置130へ送出される。データ転送処理装置130は、送られてきたリクエストの起動元プロセッサ番号番号“100”によりバッファのリードポインタ番号“000”を決定する。バッファ番号“0”～“2”からのリードデータおよびケット番号“3”のデータをフリップフロップに同時に入力させ、複数のケットのリクエストを完了さ

せる。また同時に到着ケット管理ビットの起動元プロセッサ番号“000”に対応する管理ビット2000～2020をエラー検出部に送出し、エラー検出部は、すべてのケットが到着していることを確認し、ケットが消滅している場合にはケット消失エラーを報告し、起動元プロセッサ番号により該プロセッサのみの切り離しによる出対運転を行う。

【0016】

【発明の効果】以上説明したように、本発明によれば、
10 各プロセッサからのデータ転送処理装置への複数のケット送出時、プロセッサ間で排他制御不要となり、共有メモリと拡張記憶装置との間のデータ転送でプロセッサにオーバーヘッドがなくなる。また、一連のリクエストとして送出された複数のケットの管理ができ一連のリクエスト送出中にプロセッサが故障した場合、該プロセッサのみの切り離しによる縮退運転が可能となる。また、最終ケットのバッファが不要となりハードウェア量が削減できる。

【図面の簡単な説明】

20 【図1】本発明の実施例の構成図である。

【図2】本発明の第1の実施例におけるリクエスト整列制御部の構成図である。

【図3】本発明の第2の実施例におけるリクエスト整列制御部の構成図である。

【図4】従来例の構成図である。

【図5】図5中の共有メモリの構成図である。

【図6】図5中のデータ転送処理装置の構成図である。

【図7】図5中の演算プロセッサの構成図である。

【符号の説明】

30 100～103 演算処理装置
110～111 ローカルネットワーク
120～123 共有メモリ
130～131 データ転送処理装置
140～141 拡張記憶装置
150 グローバルネットワーク
160 リクエスト発行制御部
200 リクエスト発行手段
210 パケット送出手段
220 ロック手段
40 300 ロック成功判断手段
310 ロックバイト
400 リクエスト整列手段
410 リクエスト解釈手段
420 転送制御手段
430 共有メモリアクセス手段
440 拡張記憶装置アクセス手段
450 終了通知手段
1000 起動元プロセッサ番号
1100 パケット番号
50 1200～1230 パケット保持バッファ

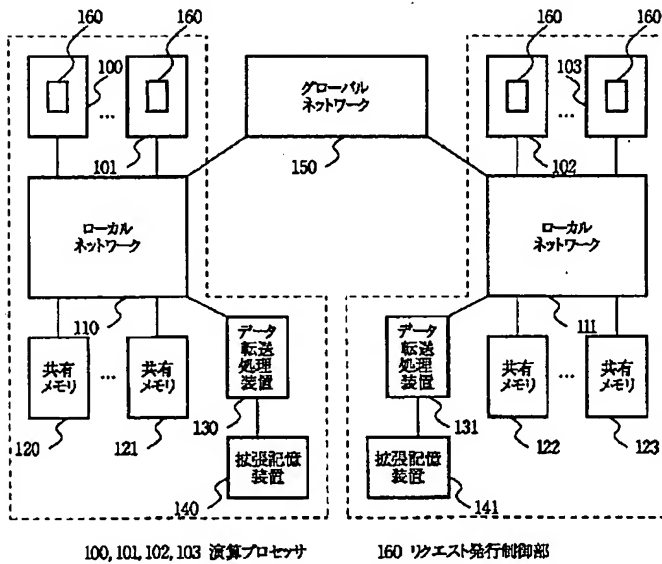
9

10

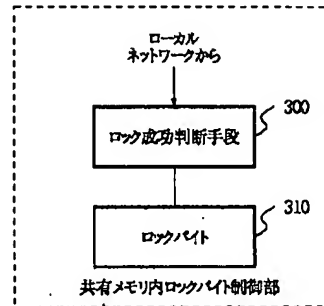
1300 リクエスト整列レジスタ
1400 パケット保持バッファ選択制御部

2000~2080 到着パケット管理ビット
2200 エラー検出部

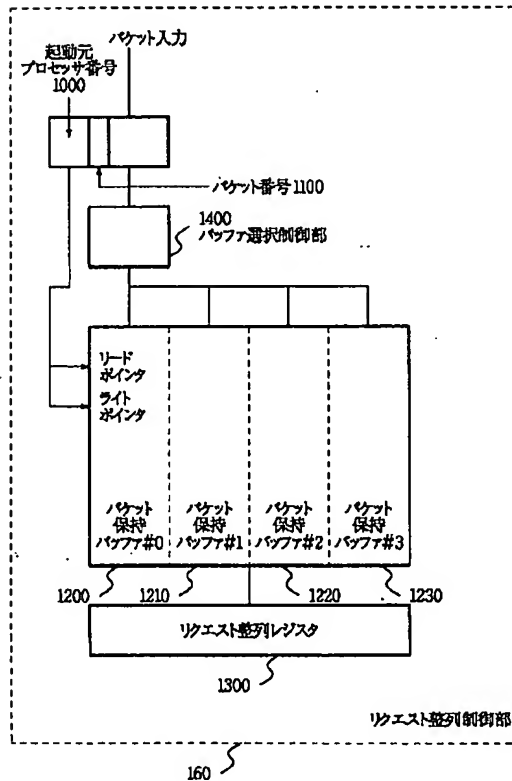
【図1】



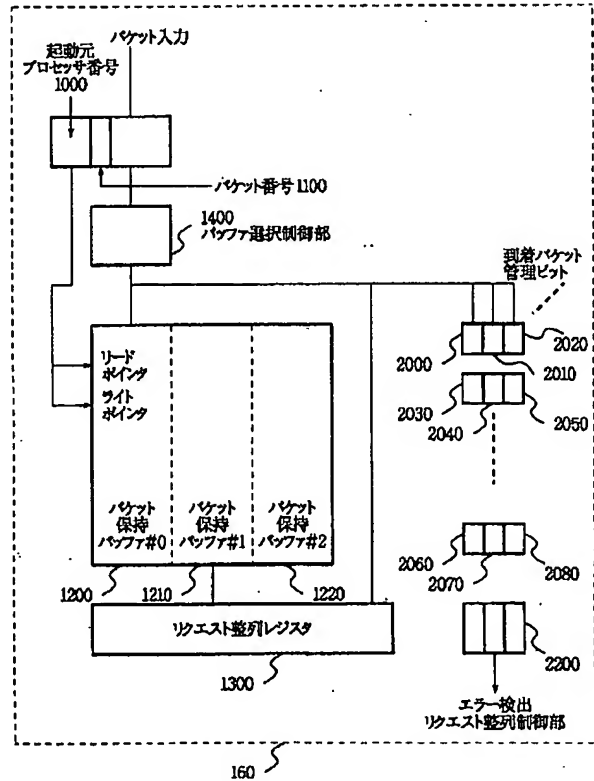
【図5】



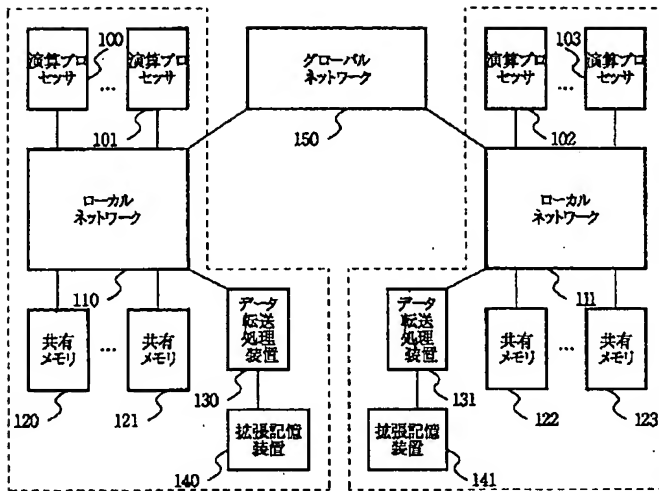
【図2】



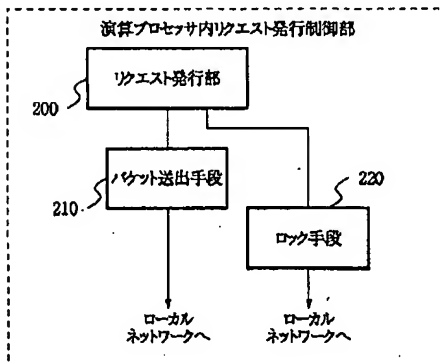
【図3】



【図4】



【図7】



【図6】

